Applications of Big Data Analytics

Mohammed M. Alani • Hissam Tawfik
Mohammed Saeed • Obinna Anya
Editors

Applications of Big Data Analytics

Trends, Issues, and Challenges

Springer

Editors
Mohammed M. Alani
Al Khawarizmi International College
Abu Dhabi, UAE

Hissam Tawfik
Leeds Beckett University
Leeds, UK

Mohammed Saeed
University of Modern Sciences
Dubai, UAE

Obinna Anya
IBM Research
San Jose, CA, USA

ISBN 978-3-319-76471-9 ISBN 978-3-319-76472-6 (eBook)
https://doi.org/10.1007/978-3-319-76472-6

Library of Congress Control Number: 2018943141

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Big Data comes in high volume, velocity, and veracity, and from myriad sources, including log files, social media, apps, IoT, text, video, image, GPS, RFID, and smart cards. The process of storing and analyzing such data exceeds the capabilities of traditional database management systems and methods, and has given rise to a wide range of new technologies, platforms, and services—referred to as Big Data Analytics. Although the potential value of Big Data is enormous, the process and applications of Big Data Analytics have raised significant concerns and challenges across scientific, social science, and business communities.

This book presents the current progress on challenges related to applications of Big Data Analytics by focusing on practical issues and concerns, such as the practical applications of predictive and prescriptive analytics especially in the health and disaster management domains, system design, reliability, energy efficiency considerations, and data management and visualization. The book is the state-of-the-art reference discussing progress made and problems encountered in applications of Big Data Analytics, as well as prompting future directions on the theories, methods, standards, and strategies necessary to improve the process and practice of Big Data Analytics.

The book comprises 10 self-contained and refereed chapters written by leading international researchers. The chapters are research-informed and written in a way that highlights the practical experience of the contributors, while remaining accessible and understandable to various audiences. The chapters provide readers with detailed analysis of existing trends for storing and analyzing Big Data, as well as the technical, scientific, and organizational challenges inherent in current approaches and systems through demonstrating and discussing real-world examples across a wide range of application areas, including healthcare, education, and disaster management. In addition, the book discusses, typically from an application-oriented perspective, advances in data science, including techniques for Big Data collection, searching, analysis, and knowledge discovery.

The book is intended for researchers, academics, data scientists, and business professionals as a valuable resource and reference for the planning, designing, and implementation of Big Data Analytics projects.

Organization of the Book

The chapters of the book are ordered such that chapters focusing on the same or similar application domain or challenge appear consecutively. Each chapter examines a particular Big Data Analytics application focusing on the trends, issues, and relevant technical challenges.

Chapter 1 discusses how recent innovations in mobile technologies and advancements in network communication domain have resulted in the emergence of smart system applications, in support of the wide range and coverage provision, low costs, and high mobility. 5G mobile network standards represent a promising cellular technology to provision the future of smart systems data traffic. Over the last few years, smart devices, such as smartphones, smart machines, and intelligent vehicles communication, have seen exponential growth over mobile networks, which resulted in the need to increase the capacity due to generating higher data rates. These mobile networks are expected to face "Big Data" related challenges, such as explosion in data traffic, storage of big data, and the future of smart devices with various Quality of Service (QoS) requirements. The chapter includes a theoretical and conceptual background on the data traffic models over different mobile network generations and the overall implications of the data size on the network carrier.

Chapter 2 explores the challenges, opportunities, and methods, required to leverage the potentiality of employing Big Data into the assessing and predicting the risk of flooding. Among the various natural calamities, flood is considered one of the most frequently occurring and catastrophic natural hazards. During flooding, crisis response teams need to take relatively quick decisions based on huge amount of incomplete and, sometimes, inaccurate information mainly coming from three major sources: people, machines, and organizations. Big Data technologies can play a major role in monitoring and determining potential risk areas of flooding in real time. This could be achieved by analyzing and processing sensor data streams coming from various sources as well as data collected from other sources such as Twitter, Facebook, satellites, and also from disaster organizations of a country by using Big Data technologies.

Chapter 3 discusses artificial intelligence methods that have been successfully applied to monitor the safety of nuclear power plants (NPPs). One major safety issue of an NPP is the loss of a coolant accident (LOCA), which is caused by the occurrence of a large break in the inlet headers (IH) of a nuclear reactor. The chapter proposes a neural network (NN) design methodology in three stages to detect the break sizes of the IHs of an NPP. The results show that the proposed methodology outperformed the MLP of the previous work. Compared with exhaustive training of

all two-hidden layer architectures, the speed of the proposed methodology is faster than that of exhaustive training. Additionally, the optimized two-hidden-layer MLP of the proposed methodology has a similar performance to exhausting training. In essence, this chapter is an example of an engineering application of predictive data analytics for which "well-tuned" neural networks are used as the primary tool.

Chapter 4 discusses a Big Data Analytics application for disaster management leveraging IoT and Big data. In this chapter, the authors propose the use of drones or Unmanned Aerial Vehicles (UAVs), in a disaster situation as access points to form an ad hoc mesh multi-UAV network that provides communication services to ground nodes. Since the UAVs are the first components to arrive at a given disaster site, finding the best positions of the UAVs is both important and non-trivial. The deployment of the UAV network and its adaption or fine-tuning to the scenario is divided into two phases. The first phase is the initial deployment, where UAVs are placed using partial knowledge of the disaster scenario. The second phase addresses the adaptation to changing conditions where UAVs move according to a local search algorithm to find positions that provide better coverage of victims. The suggested approach was evaluated under different conditions of scenarios. The number of UAVs have demonstrated a high degree of coverage of "victims."

From a Big Data Analytics perspective, the goal of the application is to determine optimum or near-optimum solutions in a potentially very large and complex search space. This is due to the high dimensionality and huge increase of parameters and combinatorics, with the increase in the number of UAVs and size and resolution of the disaster terrain. Therefore, this is considered an application of data analytics, namely prescriptive or decision analytics using computational intelligence techniques.

Chapter 5 proposes a novel health data analytics application based on deep learning for sleep apnea detection and quantification using statistical features of ECG signals. Sleep apnea is a serious sleep disorder phenomena that occurs when a person's breathing is interrupted during sleep. The most common diagnostic technique that is used to deal with sleep apnea is polysomnography (PSG), which is done at special sleeping labs. This technique is expensive and uncomfortable. The proposed method in this chapter has been developed for sleep apnea detection using machine learning and classification including deep learning. The simulation results obtained show that the newly proposed approach provides significant advantages compared to state-of-the-art methods, especially due to its noninvasive and low-cost nature.

Chapter 6 presents an analysis of the core concept of diagnostic models, exploring their advantages and drawbacks to enable initialization of a new pathway toward robust diagnostic models that overcome current challenges in headache disorders. The primary headache disorders are the most common complaints worldwide, and the socioeconomic and personal impact of headache disorders are very significant. The development of diagnostic models to aid in the diagnosis of primary headaches has become an interesting research topic. The chapter reviews trends in this field with a focus on the analysis of recent intelligent systems approaches with respect to the diagnosis of primary headache disorders.

This chapter demonstrates a novel Resource Allocation Scheme (RAS) and algorithm along with a new 5G network slicing technique based on classification and measuring the data traffic to satisfy QoS for smart systems such as smart healthcare application in a smart city environment. The chapter proposes the RAS for efficient utilization of the 5G radio resources for smart devices communication.

Chapter 7 reports on an application of Big Data analytics in education. The past decade witnessed a very significant rise in the use of electronic devices in education at all educational levels and stages. Although the use of computer networks is an inherent feature of online learning, the traditional schools and universities are also making extensive use of network-connected electronic devices such as mobile phones, tablets, and computers. Data mining and Big Data analytics can help educationalists to analyze enormous volume of data generated from the active usage of devices connected through a large network. In the context of education, these techniques are specifically referred to as Educational Data Mining (EDM) and Learning Analytics (LA). This chapter discusses major EDM and LA techniques used in handling big data in commercial and other activities and provides a detailed account of how these techniques are used to analyze the learning process of students, assessing their performance and providing them with detailed feedback in real time. The technologies can also assist in planning administrative strategies to provide quality services to all stakeholders of an educational institution. In order to meet these analytical requirements, researchers have developed easy-to-use data mining and visualization tools. The chapter discusses, through relevant case studies, some implementation of EDM and LA techniques in universities in different countries.

Chapter 8 attempts to address some of the challenges associated with Big Data management tools. It introduces a scalable MapReduce graph partitioning approach for high-degree vertices using master/slave partitioning. This partitioning makes Pregel-like systems in graph processing, scalable and insensitive to the effects of high-degree vertices while guaranteeing perfect balancing properties of communication and computation during all the stages of big graphs processing. A cost model and performance analysis are given to show the effectiveness and the scalability of authors' graph partitioning approach in large-scale systems.

Chapter 9 presents a multivariate and dynamic data representation model for the visualization of large amount of healthcare data, both historical and real-time for better population monitoring as well as for personalized health applications. Due to increased life expectancy and an aging population, a general view and understanding of people health are more urgently needed than before to help reducing expenditure in healthcare. The chapter proposes a multivariate and dynamic data representation model for the visualization of large amounts of healthcare data, both historical and real time.

Chapter 10 presents the adaptation of the big data analytics methods for software reliability assessment. The proposed method uses software with similar properties and known reliability indicators for the prediction of reliability of a new software. The concept of similar programs is formulated on the basis of five principles. Search results of similar programs are described. Analysis, visualization, and interpreting for offered reliability metrics of similar programs are executed. The

chapter concludes with reliability similarity for comparable software based on the use of metrics for prediction of new software reliability. The reliability prediction presented in this chapter aims at allowing developers to operate resources and processes of verification and refactoring potentially increasing software reliability and cutting development cost.

Abu Dhabi, UAE — Mohammed M. Alani
Leeds, UK — Hissam Tawfik
Dubai, UAE — Mohammed Saeed
San Jose, CA, USA — Obinna Anya

Contents